

Amendments to the Claims:

1. (Currently Amended) A document extracting device, comprising:

a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the candidate documents; and

a document extracting device to extract a combination of documents ~~whose~~ from the candidate documents with a sum of the degrees of similarity between the candidate documents computed by the similarity computing device is that is the smallest when any number of the combination of documents are extracted from among a group of the candidate documents.
2. (Currently Amended) The document extracting device according to Claim 1,

the similarity computing device comprising:

a character-string-dividing functional unit to divide each of the candidate documents into predetermined character strings;

a character-string frequency computing functional unit to compute document vectors of the candidate documents on the basis of ~~the a~~ a frequency of appearance of the predetermined character strings divided by the character-string-dividing functional unit; and

a mutual similarity computing functional unit to compute the degrees of similarity between the candidate documents on the basis of the document vectors obtained from the character-string frequency computing functional unit.
3. (Currently Amended) The document extracting device according to Claim 2,

the character-string-dividing functional unit dividing each of the candidate documents into predetermined character strings using any ~~one of the following~~ one of the following character string division methods: a morphological analysis method, an n-gram method, and a stop-word method.

4. (Currently Amended) The document extracting device according to Claim 2, the character-string frequency computing functional unit generating document vectors obtained by weighting each of the candidate documents by ~~TFIDF~~ a term frequency and inverse document frequency (TFIDF) weighting method on the basis of ~~the~~ a frequency of appearance of the divided character strings.

5. (Currently Amended) The document extracting device according to Claim 2, the mutual similarity computing functional unit computing ~~the~~ degrees of similarity between the candidate documents by a vector space method on the basis of the document vectors of the candidate documents.

6. (Currently Amended) A computer-readable media having a document extracting program allowing a computer to serve as:

a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the candidate documents; and

a document extracting device to extract a combination of documents ~~whose~~ from the candidate documents with a sum of the degrees of similarity between the candidate documents computed by the similarity computing device is that is the smallest when any number of the combination of documents are extracted from among a group of the candidate documents.

7. (Currently Amended) The ~~document extracting program~~ media according to Claim 6,

the similarity computing device comprising:

a character-string-dividing function to divide each of the candidate documents into predetermined character strings;

a character-string frequency computing function to compute document vectors of the candidate documents on the basis of ~~the~~ a frequency of appearance of the predetermined character strings divided by the character-string-dividing function; and

a mutual similarity computing function to compute the degrees of similarity between the candidate documents on the basis of the document vectors obtained by the character-string frequency computing function.

8. (Currently Amended) ~~A document extracting program~~ The media according to Claim 6,

the similarity computing device comprising:

a character-string-dividing function to divide each of the candidate documents into character strings using any one of character string division methods;

a character-string frequency computing function to generate document vectors obtained by weighting each of the documents by ~~TFIDF~~ a term frequency and inverse document frequency (TFIDF) weighting method on the basis of ~~the~~ a frequency of appearance of the divided character strings; and

a mutual similarity computing function to compute the degrees of similarity between the candidate documents by a vector space method on the basis of the document vectors of the candidate documents.

9. (Currently Amended) A document extracting method, comprising:

acquiring a plurality of documents to be candidates for extraction ~~are acquired~~; computing all degrees of similarity between the candidate documents ~~are computed~~; and ~~when any number of documents are extracted from among a group of the documents,~~ extracting a combination of documents ~~whose~~ from the candidate documents with a sum of the degrees of similarity between the candidate documents that is the smallest ~~is extracted~~

when any number of the combination of documents are extracted from among a group of the candidate documents.

10. (Currently Amended) The document extracting method according to Claim 9, further comprising:

dividing each of the documents ~~being divided~~ into predetermined character strings, ~~the computing a frequency of appearance of the divided character strings is computed,~~ computing document vectors of the candidate documents ~~are computed~~ on the basis of the frequency of appearance of the predetermined character strings, and then computing the degrees of similarity between the candidate documents ~~to be candidates for extraction are computed~~ using the document vectors.

11. (Currently Amended) The document extracting method according Claim 9, further comprising:

dividing each of the candidate documents ~~being divided~~ into predetermined character strings using any one of character string division methods, ~~such as including a~~ morphological analysis method, an n-gram method, and a stop-word method, computing document vectors of the candidate documents ~~obtained by~~ weighting each of the documents by ~~TFIDF~~ a term frequency and inverse document frequency (TFIDF) weighting method on the basis of ~~the a frequency of appearance of the divided predetermined character strings are computed,~~ and computing the degrees of similarity between the candidate documents ~~to be candidates for extraction are computed~~ using a vector space method on the basis of the document vectors.

12. (Currently Amended) A document extracting device, comprising:

a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the candidate documents; and

a document extracting device to extract a combination of documents based on ~~the a sum of the~~ degrees of similarity between the candidate documents ~~computed by the similarity computing device~~ when any number of the combination of documents are extracted from among a group of the candidate documents.

13. (Currently Amended) A computer-readable media having a document extracting program allowing a computer to serve as:

a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the candidate documents; and

a document extracting device to extract a combination of documents based on the a sum of the degrees of similarity between the candidate documents ~~computed by the similarity computing device~~ when any number of the combination of documents are extracted from among a group of the candidate documents.

14. (Currently Amended) A document extracting method, comprising:

acquiring a plurality of candidate documents to be candidates for extraction are ~~acquired~~;

computing all degrees of similarity between the candidate documents are ~~computed~~; and

~~when any number of documents are extracted from among a group of the documents, extracting~~ a combination of documents from the candidate documents based on ~~the a sum of the~~ degrees of similarity between the candidate documents ~~is extracted when any number of the combination of documents are extracted from among a group of the candidate documents.~~